

EXTERNALITIES AND THE REGULATORY ROLE OF INFORMATION DISCLOSURE IN MARKETS

BY JULIEN MANILI*

Abstract. Disclosure about many externality-related attributes is not mandatory, and consumers have to rely on the information strategically disclosed by manufacturers. I show that under voluntary disclosure, manufacturers can mislead consumers about externalities by exploiting a phenomenon of wishful belief formation. Bounded reasoning and inattentive consumers prevent attentive consumers from deducing production processes from disclosure incentives. The resulting subjective uncertainty gives bite to wishful belief formation, which manufacturers can exploit through a strategic use of vagueness. These misperceptions reduce the pressure of the demand side on dirty production, with important implications for the role of mandatory disclosure to regulate externalities. (*JEL* D82, D83, Q58, L15, O33)

“Shed light on a deceptive man, and he will behave as an honest man.”
—Napoléon Bonaparte.

1. INTRODUCTION

What consumers purchase plays a central role in determining how goods are produced. The available alternatives are produced with diverse production processes, and hence the purchases of consumers also constitute choices of production processes. In a context where consumers increasingly care about externalities of production, market forces could induce producers to adopt and develop cleaner technologies. Yet, in the current institutional setting, disclosure is not mandatory for most aspects of production related to externalities, and consumers have to rely on the information strategically disclosed by manufacturers to appraise the externalities of alternatives. This makes consumer choice depend on *perceptions* of externalities, rather than the externalities themselves.

Most of the literature on externalities implicitly assumes that consumer choice reflects true preferences, that is, that consumers have *correct* perceptions about the externalities associated to alternatives. This led to the traditional view of externalities as a problem of consumers’ *preferences*, who do not account enough for the collateral

Date: August 16, 2022.

*Northwestern University (email: julienmanili2028@u.northwestern.edu). I am grateful to Pierpaolo Battigalli, Emiliano Catonini, Nicolò Generoso, Nicola Gennaioli, Alfonso Maselli, Massimo Morelli, Pietro Ortoleva, Marco Ottaviani, and Alexei Verkhovtsev for very helpful comments.

effects of their choices. Growing evidence instead indicates that consumers substantially and increasingly care about externalities,¹ and Aghion et al. (2022) show that when consumers are correctly informed, an increase in consumers' environmental concern can induce green technological change in proportions superior to a large fuel price increase. The failure of markets to incorporate these preferences suggests that the issue of externalities could be in great part driven by a problem of *information*, rather than by one of preferences.² The mechanism is simple: when disclosure is not mandatory, firms using "dirty" processes can employ powerful marketing artifices to be *perceived* clean, which reduces the pressure of the demand side on dirty production and fosters firms' incentives to use dirty technologies.

This paper shows that indeed, under the regime of voluntary disclosure major informational issues about externalities arise in markets. The "dirty" manufacturers manage to mislead consumers about externalities by exploiting a phenomenon of wishful belief formation—a tendency of consumers to believe the most agreeable property given subjective uncertainty.³ This belief formation mechanism is highly likely for attributes related to externalities, as I shall argue below. On the other hand, bounded strategic reasoning and the presence of inattentive consumers prevent attentive consumers from deducing production processes by reasoning about disclosure incentives upon observing vague disclosures. The resulting subjective uncertainty gives bite to wishful belief formation and allows the dirtiest manufacturers to be perceived cleaner through a strategic use of vagueness.

The psychology of wishful belief formation is highly likely for externalities. Most features of processes related to externalities (country of production, use of pesticides, and so on) have no detectable consequences on the experience of consumption, and hence consumers never learn the truth *post facto*. The absence of resolution of uncertainty offers the ideal ground for wishful beliefs to flourish, as consumers can believe the most agreeable property consistent with their subjective uncertainty without expecting to be disappointed or adversely affected by experience.⁴ This inhibits mechanisms triggering caution in belief formation, such as disappointment aversion or danger.⁵ Recent experimental findings of Drobner (2022) show that, indeed, the

¹On environmental concern, see Aghion et al. (2022) and Andre et al. (2022); on attitudes toward social conditions of production and inequalities, see Hill (2021) and Hill & Lloyd (2021).

²In other words, consumers may not account for externalities in their choices due to underestimation of externalities rather than lack of care. The two have the same observable consequences, but suggest very different remedies.

³An expression of the phenomenon of cognitive dissonance described by Akerlof & Dickens (1982).

⁴This is not the case for attributes whose true state determines experience. If the yogurt is pineapple flavored, pineapple flavor is the experienced flavor, independently from the beliefs formed at the moment of purchase. Drobner (2022) shows that strong feedback from the true state tends to trigger *pessimistic* belief formation, the opposite of wishful belief formation.

⁵*Eyes do not see, heart does not hurt.* (Old Italian saying.)

expectation of the absence of resolution of uncertainty strongly predicts wishful belief formation.⁶ Naturally, this psychology will not escape manufacturers' marketing departments, whose industry is to identify and exploit these phenomena to maximize the perceived value of consumers. Manufacturers will thus take it into account in the design of their labels.

I formally study consumer perceptions and disclosure behavior in a disclosure model à la Grossman (1981). A manufacturer is privately informed about the attributes of his product—e.g., aspects of the production process—and discloses information on the label with the goal of maximizing the perceived value of consumers (disclosure is assumed costless). Going through the shelves, consumers observe the label and try to understand the true properties of the product. Disclosure regulations forbid misreporting, but disclosure is at the manufacturer's discretion.

The analysis of consumer perceptions proceeds in two steps. First, I analyze the subjective uncertainty of consumers by characterizing the strategic inferences on processes consumers can reach from the information disclosed on labels as a function of their degree of reasoning. This is achieved by means of a rationalizability procedure, whose finite steps capture bounded reasoning, while the limit captures infinitely sophisticated strategic reasoning. Then I predict consumer perceptions with wishful belief formation over the uncertainty consistent with information and reasoning, and by analyzing how manufacturers disclose information in reaction to this psychology.

According to the classical “unraveling result” of Grossman (1981) and Milgrom (1981), while sellers may not report all their information, consumers can deduce the sellers' private information from disclosure incentives, the higher quality sellers having incentives to distinguish themselves from the lower quality sellers.⁷ Although vagueness in disclosures is pervasive in supermarkets, the unraveling result implies that consumers have correct perceptions, even when information is withheld by the seller. Empirical and experimental evidence show instead that consumers tend to be overoptimistic about hidden information.⁸ The unraveling mechanism relies on conditions that are highly unlikely in supermarkets—among others: (i) infinitely sophisticated strategic reasoning; (ii) the absence of consumers engaging in limited reasoning; (iii) the absence of inattentive consumers. Failure of any of these assumptions makes nondisclosure consistent with strategic behavior of several types of the sellers. This leads to the possibility of rationalizing vague labels in multiple ways, giving bite to wishful belief formation and possibly inducing important misperceptions about the “dirty” processes.

⁶Drobner (2022) also shows that all experiments in the literature on motivated beliefs in which uncertainty was never resolved yielded that individuals were subject to wishful belief formation.

⁷See also Grossman & Hart (1980), Milgrom & Roberts (1986), Battigalli (2006), and Viscusi (1978).

⁸See for instance Mathios (2000); Jin, Luca & Martin (2021); Montero & Sheth (2021).

The failure of infinite strategic reasoning is specially likely for attributes related to externalities: the absence of feedback prevents strategic learning to occur.⁹ Bounded reasoning leads to a partial unraveling, where all types below a threshold quality are strategically indistinguishable; such threshold is higher for lower levels of reasoning. This is because the reasoning process leading to full unraveling proceeds as a cascade from higher to lower quality types, and so separating the lower types require a higher level of reasoning. Further, the presence of consumers engaging in limited reasoning severely impairs the ability of the sophisticated ones to draw sharp inferences on production processes. The reason is that higher steps of reasoning rely on consumers' confidence that *everybody* engages in sophisticated reasoning, and hence the presence of naive consumers impede sophisticated consumers to move forward in their reasoning. This contrasts with results of Milgrom & Roberts (1986) showing that the presence of naive consumers does not harm the information of sophisticated consumers.¹⁰

Second, I show that the presence of inattentive consumers severely limits the ability of attentive consumers to make strategic inferences, even if all attentive consumers engage in infinite strategic reasoning. Attributes lacking material feedback are not salient, which makes them specially subject to inattention. One hardly forgets to check the flavor of a yogurt before buying it, since this attribute marks experience, while we can bet that many consumers forget to reason about the origin of milk in the absence of cues recalling the existence of such attribute.¹¹ I model this by assuming that when no information is disclosed, a share of consumers is inattentive about the attribute and unconsciously perceives some *a priori* process suggested by the framing.¹² I show that nondisclosure can be rationalized as the manufacturer being of any type worse than the *a priori* of inattentive consumers; any such type may rationally find advantageous to stay silent to avoid attracting the attention of inattentive consumers on the attribute.

These sources of subjective uncertainty, when consumers are subject to wishful belief formation, allow manufacturers using the “dirtiest” processes to be perceived cleaner than they are through a strategic use of vagueness. First, to exploit bounded

⁹Jin, Luca & Martin (2021) show experimentally that in the absence of feedback, a low level of strategic reasoning in disclosure interactions is widespread and persists over repeated interactions. In contrast, they find that direct and repeated feedback eventually leads to the unraveling predictions.

¹⁰Milgrom & Roberts' result can obtain under a more demanding notion of consumer sophistication when the seller can make *partial* disclosures—see Section 6 for the details. The result certainly fails, however, when the seller can only fully disclose or stay silent, as nondisclosure becomes rationalizable as an attempt of any type of the seller but the highest to mislead naive consumers.

¹¹The role of salience in shaping recall and attention is abundantly documented. See Bordalo et al. (2022); Bordalo, Gennaioli & Schleifer (2020), and the references therein.

¹²Such *a priori* may be formed through unconscious processes influenced by the framing, such as cued recall or stereotypical judgment (see, e.g., Gennaioli & Schleifer 2010; Bordalo, Gennaioli & Shleifer 2013, 2020).

reasoning manufacturers have interest to be vague upward in their disclosures (the dirtiest one selects nondisclosure); then low levels of reasoning entail important misperceptions about dirty processes. Second, in the presence of inattentive consumers, even infinitely sophisticated consumers are subject to misperceptions; after observing nondisclosure, they perceive the highest process worse than the *a priori* of inattentive consumers. Then all manufacturers worse than such *a priori* have interest to pool into nondisclosure. When this *a priori* is highly ecological—and firms have interest to employ their genius to this effect—this may entail considerable misperceptions. A striking feature in both of these mechanisms is that all processes benefiting from misperceptions are perceived equally. This means that only the worst of these processes will survive market selection due to a local “lemons” problem.

It should be emphasized that wishful belief formation is essential to explain both consumer misperceptions and the strategic use of vagueness of manufacturers. The mechanisms creating failure of unraveling (i.e., subjective uncertainty on processes) do not alone predict misperceptions. If consumers were subject to pessimistic belief formation and would believe the worst process among those perceived possible, then any form of subjective uncertainty would penalize sellers rather than benefit them. In this case, I show that manufacturers cannot exploit vagueness and consumers have correct perceptions regardless of their level of reasoning. So, the connection between bounded reasoning and misperceptions is the effect of two subtle forces: wishful belief formation and the consequent incentives of manufacturers to be strategically vague.

These misperceptions have important implications for the ecological transition. They entail an “unraveling” of the green market, where dirty production drives out green production. In Section 8, I study this mechanism in a model of consumer choice with endogenous perceptions. The fact that clean processes are signaled to the market is not enough; the fundamental problem of optional disclosure lies in misperceptions about “dirty” processes. The dirty alternatives being cheaper and perceived cleaner than they are, they absorb the demand of green producers. Differently from Akerlof’s (1970) “lemons” problem, such unraveling can occur even if clean and dirty alternatives are *differentiated*. The reason is that clean producers lose market shares due to a *reduced* perceived difference with the dirty alternatives. Only the consumers with high willingness to pay for such reduced difference buy the green alternative.¹³ This may importantly reduce the size and profitability of the clean market, depressing green technology adoption and green R&D investments.

Mandatory disclosure of processes could thus be extremely effective to regulate externalities, by restoring the pressure of the demand side on “dirty” production. If firms have to disclose their processes *including when dirty*, the profitability of dirty processes will be severely undermined. Mathios (2000) finds evidence of a

¹³Of course, all consumers perceiving that the dirty alternative is green (due to, e.g., inattention or insufficient reasoning) buy the dirty alternative regardless of their environmental concern.

major negative demand shock on “dirty” goods after a regulatory shift from voluntary to mandatory disclosure in the salad dressing market.¹⁴ This will trigger virtuous dynamics for the green transition. The negative demand shock on “dirty” goods will drive their prices down and make them less profitable, incentivizing producers to switch on more ecological technologies. The entry of producers in clean markets will intensify competition on ecological goods, reducing their prices and production costs through a higher pressure to clean innovation. The increased affordability of ecological goods will be a further demand shock for “dirty” goods. This process being self-reinforcing, disclosure laws could create the conditions for the ecological transition. The results of Aghion et al. (2022) support these conclusions once the role of information is accounted (see Manili 2022 for the details).

Relation to the literature. This paper contributes to the literature in several ways. First, it connects the problem of externalities to the problem of asymmetric information in markets (Akerlof 1970; Spence 1977). The literature on externalities typically assumes “revealed preferences” and, through this lens, essentially views externalities as a problem of *preferences*. On the other hand, increasing evidence shows that consumers intrinsically care about externalities (Aghion et al. 2022; Andre et al. 2021; Hill & Lloyd 2021) and Aghion et al. (2022) show that these preferences (if expressed in consumer choice) induce large-scale green technology adoption. This paper reconciles this conflicting evidence by pointing out important misperceptions about externalities arising under voluntary disclosure. Mandatory disclosure is then essential to exploit the market mechanism of Aghion et al. (2022), since even if consumers have highly ecological preferences, misperceptions are likely to prevent these preferences from being reflected in consumer choice. This provides the theoretical basis for a novel, high-impact intervention to regulate externalities.

Second, this paper contributes to disclosure theory by incorporating the psychology of motivated belief formation into the analysis of disclosure.¹⁵ The basic insight is that wishful belief formation, when combined with mechanisms leading to failures of unraveling, allows “dirty” firms to mislead consumers through a strategic use of vagueness.¹⁶ I introduce a solution concept that combines conscious strategic reasoning, which determines subjective uncertainty, and unconscious belief formation over subjective uncertainty. This delivers a transparent and psychologically realistic

¹⁴See also Kim, Kim & Arora (2022), who show experimentally that mandatory disclosure of the GMO status induces a major negative demand shock on GMO products relative to optional disclosure.

¹⁵See Bénabou & Tirole (2016) for a recent account of the growing literature on motivated beliefs.

¹⁶This insight extends to other known sources of failure of unraveling in the literature, such as disclosure costs (Verecchia 1983; Jovanovic 1982); uncertainty about whether the seller is informed (Dye 1985; Farrell 1986; Jung & Kwon 1988); unawareness (Li & Schipper 2018; Heifetz, Meier & Schipper 2020); heterogeneous preferences of consumers (Koessler & Renault 2012; Bond & Zeng 2022).

non-equilibrium analysis of belief formation.¹⁷ The mechanisms creating subjective uncertainty on processes do not alone predict misperceptions. In most models of the literature on failures of unraveling, there is multiplicity of equilibria and often full unraveling is actually an equilibrium. The solution concept that I put forward yields unique predictions as a function of the degree of reasoning and the hypothesized unconscious belief formation mechanism. Compared to previous literature, this allows to understand for which attributes issues of misperceptions are most likely to arise on the basis of psychological evidence.

A widely held view is that sophisticated consumers have correct perceptions and that misperceptions are the result of limited strategic sophistication (e.g., Milgrom & Roberts 1986; Li & Schipper 2020). This paper shows that the psychology of belief formation is even more essential. First, under pessimistic belief formation consumers have correct perceptions *regardless of their sophistication*. Second, even sophisticated consumers are likely to be subject to severe misperceptions under wishful belief formation. I find that the presence of naive consumers may severely impede sophisticated consumers to draw sharp strategic inferences, in contrast with known results of Milgrom & Roberts (1986). The presence of inattentive consumers have similar adverse effects. I find a pooling similar to the one of Hirshleifer, Lim & Teoh (2004), where all types worse than the *a priori* of inattentive consumers pool into nondisclosure.¹⁸ These results reconcile the theory in settings of costless disclosure (Grossman 1981; Milgrom 1981) with positive evidence of a persistent impossibility to deduce processes in supermarkets, in spite of applying the wisest judgment.

The rest of the paper is structured as follows: Section 2 is a thought experiment to build intuition on the model and analysis. Section 3 introduces the disclosure model. Section 4 introduces the solution concept. Section 5 goes through the unraveling result. Section 6 studies bounded reasoning and its implications. Section 7 studies the role of the presence of inattentive consumers. Section 8 showcases the implications of the results on misperceptions for the green market in a model of consumer choice. Section 9 concludes. The Appendix contains all the proofs.

2. A THOUGHT EXPERIMENT

This section presents two real-world cases of disclosure I collected in a supermarket to build intuition on the phenomena that I will analyze and on how strategic reasoning allows (or not) to make inferences from disclosure incentives.

¹⁷It is well-documented in experiments that non-equilibrium analysis provides more accurate predictions than equilibrium analysis in disclosure interactions; see, e.g., Hagenbach & Perez-Richet (2018); Li & Schipper (2020); Jin, Luca & Martin (2021); Montero & Sheth (2021).

¹⁸Differently from Hirshleifer, Lim & Teoh (2004), in my analysis the pooling is uniquely predicted (there are many other equilibria) and strategically founded—it is due to manufacturers taking advantage of the psychology of wishful belief formation. Due to the very same psychology, the pooling also entails worse misperceptions.

Suppose we are in front of the shelves in a supermarket, in France, and we wish to choose some honey jar. In particular, examine one that says:

“Honey put in the jar in the Pyrénées” (French mountains)
There is a French flag (Cultivated in France)

At first glance, it seems a French honey *cultivated in the mountains*. But it is not stated “Honey *cultivated* and put in the jar in Pyrénées”. We can infer that this honey was *not* cultivated in the Pyrénées, since if it were true, the producer would have hastened to state such information on the label. Our inference is thus unambiguous: French honey *not* cultivated in the mountains. Yet, consumers engaging in insufficient strategic reasoning will judge that a mountain seller could disclose such label, and hence, by wishful belief formation, will believe that the honey was indeed cultivated in the mountains.

Now, consider another honey jar. The label of this one says:

“Organic honey”

We can conclude that this honey was *not* cultivated in France, nor in the mountains, since all French consumers adore such properties.¹⁹ However, we cannot infer where it was cultivated precisely, it could be from Spain or China, for instance. Indeed, both are consistent with strategic behavior of the manufacturer. Both may target with such label consumers that, by inattention, do not reason about this attribute and unconsciously appraise the honey as if it was from France. If the honey is from China, the seller will further inflate the perceived value of those engaging in strategic reasoning and prefer Spain over China, who will be induced to believe Spain since consistent with strategic behavior of the seller. So we cannot know the true origin of this honey, and we can expect it to benefit important misperceptions.

The phenomenon to observe is that, when facing vague labels (nondisclosure about the location of production), bounded reasoning or the presence of inattentive consumers impede to make sharp inferences, as the label can be rationalized in multiple ways, giving bite to wishful belief formation. Moreover, it is the firms using the “dirtiest” processes that benefit the most from the resulting misperceptions. The analysis of this paper will formalize these intuitions.

We were particularly attentive about the location of production because it is a key statistic for judging environmental externalities, as it is related to pollution and agricultural methods. Leaving aside agricultural standards, farmers whose production is intended for exportation typically use more aggressive and environmentally damaging methods to reduce their unit cost and export in nations that are themselves able to produce the commodity. The relocation of productions is an essential aspect of the ecological transition.

¹⁹A preference for local products is widespread among consumers, for motives not exclusively related to ecology. Consumers’ preferences (hence, disclosure incentives) are inherently location dependent.

The disclosure game should be interpreted as played *attribute-wise*. A product should be thought of as a bundle of attributes, including those aspects of processes that do not necessarily affect its material quality—such as the country of production of ingredients, the use of genetically modified organisms, of certain harmful pesticides, and so on. For each attribute there are different possible grades over which consumers have preferences, and the disclosure game is interpreted as relative to a given attribute: the consumer looks for the disclosed information related to that attribute and tries to infer its grade.

3. A SIMPLE MODEL OF DISCLOSURE

Consider the disclosure interaction between a seller and consumers in a supermarket.²⁰ The seller is privately informed about the grade θ of a feature of his product (called the seller's *type*). Consumers only know that $\theta \in \Theta = \{\theta^1, \dots, \theta^K\}$ and have preferences over the possible grades represented by a value function $v(\cdot) : \Theta \rightarrow \mathbb{R}$ such that $v(\theta^1) < \dots < v(\theta^K)$. In our application to externalities, this means that externality-concerned consumers subjectively deem process θ^K as the “most virtuous” one, while θ^1 is deemed the “least virtuous” one. With this, it is convenient to order processes according to consumers' preference, namely: $\theta^1 < \dots < \theta^K$.

The seller can disclose messages about his type on the label of the form $m \subseteq \Theta$, interpreted as the statement “ $\theta \in m$ ”. For oranges, we may have as possible grades $\Theta = \{\text{Spain, Puglia, Sicily}\}$. The statement “Italian orange” is $m = \{\text{Puglia, Sicily}\}$; $m = \Theta$ corresponds to no disclosure of information; while $m = \{\text{Sicily}\}$ is the full disclosure statement “Sicilian orange.” Let $M = \{m \subseteq \Theta : m \neq \emptyset\}$ be the set of messages that the seller may disclose.

The consumers, going through the shelves, observe information m disclosed about θ and estimate the value of the product. For our purposes of studying disclosure behavior and consumer perceptions, we need not model explicitly price formation and purchasing decisions. The essence of incentives in this interaction lies exclusively in two forces. First, consumers examine the information disclosed about an attribute with the goal of understanding its true value. We can capture these preferences by assuming that consumers have a disutility from misperceiving the true value: a consumer's payoff of perceiving value $\hat{v} \in V := \mathbb{R}$ for a product of type θ is

$$(3.1) \quad u_2(\theta, m, \hat{v}) = -(\hat{v} - v(\theta))^2.$$

Second, the seller's goal when disclosing information is to maximize the perceived value of consumers. This is captured by assuming that the seller's payoff is a strictly increasing function $f(\cdot)$ of consumers' perceived value. This way we accommodate situations where the seller wants to maximize consumers' willingness to pay (Grossman 1981), sales (Milgrom 1981; Milgrom & Roberts 1986), or market price (Dye 1985; Verrecchia 1983; Jovanovic 1982)—variables that depend positively on the

²⁰The model builds on Grossman (1981) and Battigalli (2006).

perceived value of consumers. More generally, the seller need not have a precise model of the demand side to have these preferences, it is enough that he realizes that his business is better off, all else equal, when consumers perceive a higher value of his product, since it favors the likelihood of purchase against the competing alternatives at any given prices. Furthermore, disclosure regulations forbid misreporting, although positive disclosure is not required. If the seller makes a false statement about his type, that is, if $\theta \notin m$, he has to pay a large fine $P > \sup f$ (assume that f is bounded). These laws ensure that consumers can believe that the information disclosed by manufacturers is truthful, since it is not rational for sellers to misreport. To sum up, the seller's payoff is given by

$$(3.2) \quad u_1(\theta, m, \hat{v}) = \begin{cases} f(\hat{v}) & \text{if } \theta \in m, \\ f(\hat{v}) - P & \text{if } \theta \notin m. \end{cases}$$

A strategy $\hat{v}(\cdot) : M \rightarrow V$ for the consumer describes how she perceives her value depending on the information $m \in M$ disclosed. Let $S_2 = V^M$ be the set of strategies for the consumer.

3.1. Beliefs and best responses. Players form (probabilistic) beliefs about their co-player's type and behavior as the play unfolds.

The seller's beliefs are represented by a probability measure over the strategies of consumers, viz. $\mu^1 \in \Delta(S_2)$.²¹

The consumers' beliefs are represented by a system of conditional beliefs $(\mu^2(\cdot|m))_{m \in M} \in [\Delta(\Theta)]^M$ about the seller's type. This describes the belief $\mu^2(\cdot|m) \in \Delta(\Theta)$ that the consumer would form about the seller's type after observing each disclosure m .²²

For the seller, rationality consists in disclosing a message that maximizes his subjective expected utility. Formally, message $m \in M$ is a *best response* for type θ under beliefs $\mu^1 \in \Delta(S_2)$, written $m \in BR_1(\theta, \mu^1)$, if m maximizes the seller's expected payoff under μ^1 .

For consumers, rationality is described by sequential optimality of strategies: the consumer best responds to her beliefs conditional on each disclosure. From (3.1), for each $p \in \Delta(\Theta)$ the unique best response to p is the expected value of $v(\cdot)$ given p :

$$(3.3) \quad BR_2(p) := \arg \max_{\hat{v} \in V} \mathbb{E}_p[-(\hat{v} - v(\cdot))^2] = \mathbb{E}_p[v(\cdot)].$$

A strategy $\hat{v} \in S_2$ is a *sequential best response* to $\mu^2 \in [\Delta(\Theta)]^M$, written $\hat{v} \in BR_2^*(\mu^2)$, if $\hat{v}(m) = BR_2(\mu^2(\cdot|m))$ for every $m \in M$.

²¹For any topological space X , $\Delta(X)$ denotes the set of Borel probability measures on X .

²²One could also represent consumers' beliefs by including a description of prior beliefs $\mu^2(\cdot) \in \Delta(M \times \Theta)$ about type-dependent disclosure behavior, and require conditional beliefs to be obtained via Bayes' rule whenever possible. Yet, initial beliefs play no role in our analysis and the results are invariant to this richer representation of beliefs.

4. STRATEGIC REASONING AND BELIEF FORMATION

In this section, I introduce a solution concept to analyze disclosure and consumer perceptions. I model the belief formation process of consumers in two stages. First, I model what consumers infer about the seller's type after observing any disclosure as a function of the degree of strategic reasoning in which they engage. After reasoning about disclosure m , consumers form beliefs over the remaining subjective uncertainty according to wishful belief formation. The seller anticipates this psychology and discloses strategically in response to it.

4.1. Strategic reasoning. Consumers' strategic reasoning is modeled with Strong Δ -rationalizability (Battigalli 2003; Battigalli & Siniscalchi 2003), a rationalizability procedure for sequential games that captures forward-induction reasoning, as formalized by "best rationalization" (Battigalli 1996): when consumers observe a label, they try to rationalize it by ascribing to the manufacturer the highest degree of strategic sophistication compatible with the label.

The symbol $\Delta := (\Delta_1, \Delta_2)$ denotes restricted sets of the players' beliefs, allowing to account for how transparent features of beliefs shape strategic thinking and belief formation. The model of strategic thinking of consumers is as follows: no explicit assumptions on the seller's beliefs are made, viz. $\Delta_1 := \Delta(S_2)$, while it is assumed commonly understood that consumers feature a mild form of skepticism toward withheld information: after observing disclosure m , consumers have at least a slight suspicion (i.e., assign non-zero probability) that the seller is the worst type in m :

$$\Delta_2 := \{\mu^2 \in [\Delta(\Theta)]^M : \forall m \in M, \mu^2(\min m|m) > 0\}.$$

I shall refer to this restriction as *mild skepticism*, as in Battigalli (2006).

The notion of strategic reasoning is then as follows: If one assumes that a player, say the seller, is rational and has beliefs in Δ_1 , then if he is of type θ , he may disclose any m that can be justified by *some* belief $\mu^1 \in \Delta_1$ as a best reply for θ . Thus, the type-dependent behavioral implications of rationality for the seller given Δ are captured by the set

$$\Sigma_1^{\Delta,1} = \{(\theta, m) : \exists \mu^1 \in \Delta_1, m \in BR_1(\theta, \mu^1)\}.$$

For each disclosure m , the set of seller's types for which m is consistent with rationality is:

$$\Theta^{\Delta,1}(m) = \{\theta \in \Theta : (\theta, m) \in \Sigma_1^{\Delta,1}\}.$$

The consumer *strongly believes* an event (e.g., the seller's rationality), if she is certain of that event unless observing evidence contradicting it (Battigalli & Siniscalchi 2002). Then if the consumer strongly believes that the seller is rational, upon observing a disclosure m consistent with rationality of the seller (that is, $\Theta^{\Delta,1}(m) \neq \emptyset$) she continues to believe in the seller's rationality and so infers from m that the seller must

be of some type $\theta \in \Theta^{\Delta,1}(m)$, so that $\mu^2(\Theta^{\Delta,1}(m)|m) = 1$. This first step of reasoning captures the “rationalization principle” (Battigalli 1996): A player should always try to interpret her information about the behavior of her opponents assuming that they are not implementing “irrational” strategies. Thus strong belief in the seller’s rationality shapes consumers’ conditional beliefs, which may refine the strategies consumers may play:

$$S_2^{\Delta,2} = \{\hat{v} \in S_2 : \exists \mu^2 \in \Delta_2, \forall m \in M, \Theta^{\Delta,1}(m) \neq \emptyset \Rightarrow \mu^2(\Theta^{\Delta,1}(m)|m) = 1, \hat{v} \in BR_2^*(\mu^2)\}.$$

Then if the seller believes that consumers are rational and strongly believe in his rationality, he shall assign probability one to $S_2^{\Delta,2}$ and best respond to some belief reflecting such expectation. This leads to a weakly smaller set of possible type-disclosure pairs $\Sigma_1^{\Delta,3}$ and a weakly smaller set of types $\Theta^{\Delta,3}(m)$ for which disclosure m might be optimal. The set $\Theta^{\Delta,3}(m)$ describes the ways m can be rationalized when assuming that the seller believes that consumers are rational and strongly believe in his rationality (such set may be empty).

This third step allows to clarify the notion of “best rationalization”. It might be that m contradicts the seller’s strategic sophistication, i.e., $\Theta^{\Delta,3}(m) = \emptyset$. In this case, strong belief in the seller’s rationality *and* strategic sophistication leaves consumers’ beliefs unrestricted after observing m . Then consumers may assign positive probability to types for which m is “irrational” even if m is consistent with the seller’s rationality (i.e., even if $\Theta^{\Delta,1}(m) \neq \emptyset$). This violates the rationalization principle. So, letting $\Theta^{\Delta,0}(m) = \Theta$ and $k(m) = \max\{k \leq 2 : \Theta^{\Delta,k}(m) \neq \emptyset\}$, (level-2) “best rationalization” requires consumers to assign probability one to $\Theta^{\Delta,k(m)}(m)$.

Continuing further this process, we capture higher degrees of mutual strong belief in rationality. The limit of the process captures the first-order belief (and behavioral) implications of “common *strong* belief in rationality” (Battigalli & Siniscalchi 2002).²³

Definition 1. Let $\Sigma_1^{\Delta,0} := \Theta \times M$ and $S_2^{\Delta,0} := S_2$ and define recursively for $k \in \mathbb{N}_0$,

$$\Sigma_1^{\Delta,k+1} := \{(\theta, m) \in \Theta \times M : \exists \mu^1 \in \Delta_1, \mu^1(S_2^{\Delta,k}) = 1, m \in BR_1(\theta, \mu^1)\}.$$

For each m , define $\Theta^{\Delta,k}(m) := \{\theta \in \Theta : (\theta, m) \in \Sigma_1^{\Delta,k}\}$, let $k(m) := \max\{\ell \leq k : \Theta^{\Delta,\ell}(m) \neq \emptyset\}$, and let $\hat{\Theta}^{\Delta,k}(m) := \Theta^{\Delta,k(m)}(m)$. With this, define:

$$S_2^{\Delta,k+1} := \{\hat{v} \in S_2 : \exists \mu^2 \in \Delta_2, \forall m \in M, \mu^2(\hat{\Theta}^{\Delta,k}(m)|m) = 1, \hat{v} \in BR_2^*(\mu^2)\}.$$

The elements of $\Sigma_1^{\Delta,\infty} := \bigcap_{k \in \mathbb{N}} \Sigma_1^{\Delta,k}$ and $S_2^{\Delta,\infty} := \bigcap_{k \in \mathbb{N}} S_2^{\Delta,k}$ are Δ -rationalizable.

²³On the foundations of strategic reasoning with restrictions on beliefs, see Battigalli & Siniscalchi (2007) and Battigalli & Prestipino (2013).

The set $\Theta^{\Delta,k}(m)$ is the set of seller's types for which label m is (Δ, k) -rationalizable; the higher is k , the higher the assumed strategic sophistication of the seller. The restriction $\mu^2(\hat{\Theta}^{\Delta,k}(m)|m) = 1$ in $S_2^{\Delta,k+1}$ captures (*level- k best rationalization*): after observing m the consumer infers $\theta \in \hat{\Theta}^{\Delta,k}(m)$, i.e., interprets m assuming the highest degree of strategic sophistication $k(m) \leq k$ of the seller consistent with m . A label may be more sophisticated for manufacturers selling θ than for those selling θ' ; in this case, sophisticated consumers infer that the type of the product is *not* θ' . This may allow consumers to understand properties also from what is *not* stated: If it is not explicitly stated that an orange is from Sicily, conclude that this orange is *not* from Sicily, otherwise the seller would be foolish since Sicilian oranges are our most preferred oranges.

4.2. Belief formation. Strategic reasoning, even if infinitely sophisticated, often does not suffice to eliminate all uncertainty. Consumers may find out that several types of sellers might reasonably disclose the same label, and logic can be of no help to further inform beliefs. It is at this stage that unconscious belief formation phenomena and subjective attitudes toward uncertainty intervene in the belief formation process.

The phenomenon of interest to us is wishful belief formation. To briefly recall, this belief formation mechanism is likely to flourish when consumers expect no feedback from the true state. Without feedback to correct perceptions, the beliefs consumers form entirely determine their experience. Then in an unconscious pursuit of agreeable experiences, consumers are induced to form the most agreeable beliefs consistent with their subjective uncertainty.

Formally, when observing a label m , if the consumer assumes k degrees of mutual strong belief in rationality, her subjective uncertainty about the seller's type is described by the set $\hat{\Theta}^{\Delta,k}(m)$. With this, level- k wishful belief formation translates as follows:

$$\hat{\Delta}_2^{(k)} := \left\{ \mu^2 \in [\Delta(\Theta)]^M : \forall m \in M, \mu^2 \left(\arg \max_{\theta \in \hat{\Theta}^{\Delta,k}(m)} v(\theta) | m \right) = 1 \right\}.$$

In words, after observing each label m , consumers form a belief among those consistent with their reasoning that maximizes their utility. The level- k wishful strategies of consumers are then given by

$$S_2^{\hat{\Delta},(k)} := \{ \hat{v} \in S_2 : \exists \mu^2 \in \hat{\Delta}_2^{(k)}, \hat{v} \in BR_2^*(\mu^2) \}.$$

Notice that there is only one wishful belief system consistent with level- k reasoning, meaning that there is a *unique* level- k wishful strategy \hat{v}^k .

Notice that I do not assume that $\mu^2 \in \hat{\Delta}_2^{(k)}$ satisfy mild skepticism. The reason is that mild skepticism is only a device consumers use to reason strategically about disclosure incentives. This does not mean that mild skepticism corresponds to how consumers *actually* form their beliefs once they finished their reasoning and inferred

$\theta \in \Theta^{\Delta,k}(m)$. Certainly, wishful belief formation is a genie exercising a superior influence post rationalization, where unconscious belief formation processes prevail over the conscious ones.²⁴

The uncertainty of consumers weakly decreases in the level of reasoning, leaving less bite to wishful belief formation (indeed, $\hat{\Theta}^{\Delta,k}(m)$ is weakly decreasing in k). It is important to highlight the dual interpretation of the level k of reasoning. First, it can represent consumers engaging in limited strategic reasoning. Second, it may represent the reasoning of sophisticated consumers who lack confidence in a high degree of mutual strong belief in rationality in their strategic environment (which is likely in supermarkets, where many consumers may feature limited strategic sophistication). Indeed, if consumers perceive likely that there are consumers who do not assume k degrees of mutual strong belief in rationality, they are forced to stop the reasoning procedure at step k because the successive step relies on such k degrees of mutual strong certainty of rationality.

The correspondence $m \mapsto \hat{\Theta}^{\Delta,\infty}(m)$ describes the subjective uncertainty of consumers when they engage and infinite reasoning and feel confident that everybody engages in infinitely sophisticated reasoning. It should be pointed out that sophisticated reasoning and wishful belief formation are not at odds. Wishful belief formation is the result of an unconscious process, not of cognitive limitations. Actually, experimental evidence shows that more sophisticated individuals display a *higher* propensity to motivated belief formation; see Kahan (2013) and Kahan et al. (2017).

In order to disentangle the role of wishful belief formation, it will be useful to compare with what happens when consumers feature pessimistic belief formation and expect the worst when facing subjective uncertainty. This is a likely belief formation for attributes from which consumers expect solid consequences in their consumption experience (e.g., flavour of a yogurt, presence of gluten, solidity, etc.). The expected resolution of uncertainty tends to trigger pessimistic belief formation, as shown experimentally by Drobner (2022). This may be explained by the activation of phenomena such as disappointment aversion, or caution in the presence of danger. Level- k pessimistic belief formation yields the following conditional beliefs:

$$\bar{\Delta}_2^{(k)} := \left\{ \mu^2 \in [\Delta(\Theta)]^M : \forall m \in M, \mu^2 \left(\arg \min_{\theta \in \hat{\Theta}^{\Delta,k}(m)} v(\theta) | m \right) = 1 \right\}.$$

Notice that lower levels of reasoning expand subjective uncertainty and thus induce more pessimistic beliefs. The exact contrary happens with wishful belief formation. Subjective uncertainty will not benefit the seller as in the case of wishful belief formation, but harm the seller, which will create incentives to avoid disclosures creating (downward) ambiguity.

²⁴The analysis could be performed with ε -skepticism where $\mu^2(\min m|m) \geq \varepsilon$ for each m and defining wishful belief formation as consumers forming an ε -skeptical belief allowed by reasoning that maximizes expected utility after every m . Letting $\varepsilon \rightarrow 0$, the results would not change.

Naturally, these phenomena will not escape manufacturers' marketing departments, whose industry is to identify the genes influencing consumers and to take them into account to disclose information in the most favorable way. The seller will thus take them into account in the design of his label (I defer the details to Section 6).

5. BENCHMARK: THE UNRAVELING RESULT

I begin with the unraveling result of Grossman (1981) and Milgrom (1981), a useful preliminary to understand the logic that allows consumers to deduce processes from disclosure incentives, and discuss some of the assumptions upon which depends the result.

The result relies on the assumption that all consumers agree in their ordinal preferences over production processes. This assumption is implicitly incorporated in our model since all consumers have preferences represented by $v(\cdot)$.²⁵ This is a reasonable assumption for attributes over which there is an obvious preference among consumers, such as certain aspects of material quality, or, in the case of externalities, if there is an obvious way to order the grades of the attribute in terms of the amount of externalities.²⁶

In this situation, mild skepticism combined with best rationalization allow consumers to deduce the type of the seller from disclosure behavior (as first shown by Battigalli 2006).

The intuition is simple, and may be illustrated in the orange example. In Italian supermarkets, consumers adore Sicilian oranges, which are known for their unique taste, and prefer to buy the local, Italian oranges to the Spanish ones for ecological reasons. Thus, Spain < Puglia < Sicily. Then the Sicilian orange sellers have interest to fully disclose to avoid giving bite to consumers' skepticism. So, if it is not stated explicitly that an orange is from Sicily, sophisticated consumers conclude that this orange is *not* from Sicily. Then the Pugliese orange sellers cannot hope to convince sophisticated consumers that they are the Sicilian type with a truthful label, and thus have interest not to understate their type to avoid being mixed with the Spanish oranges. So, Pugliese sellers disclose a message m such that Spain $\notin m$.²⁷ Thus the Spanish orange sellers cannot induce sophisticated consumers to believe a higher type by stating $\theta \in \{\text{Spain, Puglia, Sicily}\}$, as consumers infer $\theta = \text{Spain}$. Full unraveling obtains.

²⁵The results of this paper depend only on the assumption that all consumers agree in their *ordinal* preferences over Θ . When there is heterogeneity in ordinal preferences, even worse pathologies of misperceptions arise.

²⁶This assumption is likely to fail for attributes such as the flavor of a yogurt, where we can expect diversity of preferences in the population, or for some production processes, for which consumers may value differently (or be differently aware of) the externalities associated to processes.

²⁷ $\Sigma_1^{\Delta,5}(\text{Puglia}) = \{\{\text{Puglia, Sicily}\}, \{\text{Puglia}\}\}$. By best rationalization, $\Theta^{\Delta,5}(\text{"Italian orange"}) = \{\text{Puglia}\}$.

Proposition 1. *For each seller's type θ , the set of Δ -rationalizable labels are those where $\min m = \theta$ (so, $m \mapsto \Theta^{\Delta, \infty}(m) = \{\min m\}$), and the unique Δ -rationalizable strategy of consumers is $m \mapsto \hat{v}(m) = v(\min m)$.*

This result suggests that a law forbidding misreporting is sufficient for consumers to have all information: although sellers may not fully disclose their private information, strategic reasoning allows consumers to deduce processes from disclosure incentives since the higher types always have interest to distinguish themselves from the lower types.

Yet, this conclusion depends on assumptions that are unlikely in our context. First, the elimination of subjective uncertainty relies on that consumers engage in infinitely sophisticated reasoning and in their confidence that *everybody* does so. On the one hand, consumers engaging in bounded reasoning will face subjective uncertainty (and the more so the lower the level of reasoning); on the other hand, the presence of naive consumers may prevent sophisticated consumers from moving forward in their reasoning process, or even alter the strategic incentives of manufacturers. These issues are analyzed in the next section. Second, I show in Section 7 that in the presence of inattentive consumers, strong failures of unraveling occur even when all attentive consumers engage in infinitely sophisticated reasoning.

The resulting subjective uncertainty of consumers gives bite to wishful belief formation and scope for dirty manufacturers to mislead consumers. In order to highlight the essential role of wishful belief formation in these results about consumer misperceptions, I shall also show that these failures of unraveling create no issues of misperceptions under pessimistic belief formation.

6. BOUNDED REASONING AND PERCEPTIONS

In this section, I analyze the implications of limited reasoning on labeling behavior of manufacturers and consumer perceptions.

Positive evidence suggests that a low level of reasoning is widespread in supermarkets, and consumers often fail to reason on manufacturers' disclosure incentives. This contrasts with experimental evidence of Li & Schipper (2020), who find relatively high levels of reasoning in a simple disclosure game. In the context of shopping, consumers are often not even aware that they are participating to a disclosure interaction, less salient than in laboratory, and tend to take labeling at face value. Further, firms often display irrelevant statements that closely resemble but do not logically imply green properties (the mountain label of Section 2 is an example), making strategic disclosure more subtle than in laboratory and requiring a higher level of cognition to achieve strategic inferences.²⁸ This means that consumers often

²⁸See Jin, Luca & Martin (2022) for experimental evidence on the role of complex disclosure. A greater complexity tends to reduce the level of reasoning, see Alaoui & Penta (2016).

do not rule out some types from what is *not* stated when strategic reasoning would allow to do so.

When a consumer engages in level- k reasoning, her subjective uncertainty is described by $m \mapsto \Theta^{\Delta,k}(m)$, which, under wishful belief formation, yields the following information-dependent perception of the value of the object:

$$m \mapsto \hat{v}^k(m) = v(\max \Theta^{\Delta,k}(m)).$$

Given that for each m , the set $\Theta^{\Delta,k}(m)$ gets larger for lower levels of reasoning, the lower the level of reasoning the greater the overestimation of utility. For example, for $k = 1$, $\Theta^{\Delta,1}(m) = m$. So, while $\hat{v}^\infty(m) = v(\min m)$, the consumers engaging in level-1 reasoning perceive the highest property in m , i.e., $\hat{v}^1(m) = v(\max m)$.

If the seller anticipates that consumers are subject to wishful belief formation but is uncertain about the level of reasoning in which consumers engage when doing the shopping, the uncertainty of the seller on consumers' perception strategies is given by:

$$\hat{S}_2^{\hat{\Delta},*} = \cup_{k=1}^{\infty} \{\hat{v}^k\}.$$

The seller will then select a disclosure that best responds to beliefs assigning probability one to this event. Further, I shall assume that the seller is *cautious*, in the sense that he does not exclude any level of reasoning of consumers:

$$\hat{\Delta}_1 = \{\mu^1 \in \Delta(S_2) : \text{Supp } \mu^1 = \hat{S}_2^{\hat{\Delta},*}\}.$$

Disclosure m is $\hat{\Delta}$ -rationalizable for θ if m is a best reply for θ to some belief $\mu^1 \in \hat{\Delta}_1$. According to Pearce (1984, Lemma 4), m is $\hat{\Delta}$ -rationalizable for θ if and only if m is not weakly dominated for θ given uncertainty $\hat{S}_2^{\hat{\Delta},*}$.²⁹ With this, we obtain a first result on the seller's strategic use of vagueness and on consumer misperceptions on processes as a function of the degree of reasoning.

Theorem 1. *When consumers are subject to wishful belief formation, the following happens:*

- (I) *For every θ , the unique $\hat{\Delta}$ -rationalizable disclosure is $m^*(\theta) = [\theta, \theta^K] := \{\theta, \dots, \theta^K\}$ (which is weakly dominant for θ given $\hat{S}_2^{\hat{\Delta},*}$).*
- (II) *The subjective uncertainty of consumers about θ as a function of reasoning is for all $n \geq 0$: $\Theta^{\Delta,2n+1}(m^*(\theta)) = \{\theta, \dots, \theta^{K-n}\}$.*
- (III) *The consumer perception $\hat{\pi}^k(\theta) := \max \Theta^{\Delta,k}(m^*(\theta))$ of process θ as a function of the level k of reasoning is for all $n \geq 0$, $\hat{\pi}^{2n+1}(\theta) = \max\{\theta, \theta^{K-n}\}$.*

²⁹Recall that m weakly dominates m' for θ given $\hat{S}_2^{\hat{\Delta},*}$ if (i) for all $\hat{v} \in \hat{S}_2^{\hat{\Delta},*}$ we have $u_1(\theta, m, \hat{v}(m)) \geq u_1(\theta, m', \hat{v}(m'))$ and (ii) there exists $\hat{v}^* \in \hat{S}_2^{\hat{\Delta},*}$ such that $u_1(\theta, m, \hat{v}^*(m)) > u_1(\theta, m', \hat{v}^*(m'))$.

The key disclosure implication of bounded reasoning and wishful belief formation is that sellers have interest to use vagueness strategically to optimally mislead the consumers engaging in bounded reasoning. Indeed, $m^*(\theta)$ is the unique label that maximizes the effect of wishful belief formation *whatever the level of reasoning of consumers*.

The argument showing (I) is as follows: First, for every k and truthful disclosure m for θ (i.e., $\theta \in m$), we have $\Theta^{\Delta,k}(m) \cap [\theta, \theta^K] \subseteq \Theta^{\Delta,k}(m^*(\theta))$, so that $\max \Theta^{\Delta,k}(m) \leq \max \Theta^{\Delta,k}(m^*(\theta))$. This means that $m^*(\theta)$ does weakly better for θ than all other disclosures against wishful strategies. Moreover, if m does not contain some higher type $\theta' > \theta$, then there exists a level of reasoning k such that $\max \Theta^{\Delta,k}(m^*(\theta)) = \theta'$ and $\max \Theta^{\Delta,k}(m) < \theta'$, yielding a strictly lower payoff against the level- k wishful strategy $\hat{v}^k(\cdot)$. If m contains some $\theta' < \theta$, then $\min m < \theta$ and so for $k = \infty$ we have $\hat{v}^\infty(m^*(\theta)) = v(\theta) > v(\min m) = \hat{v}^\infty(m)$.

This prediction on the strategic use of vagueness is consistent with positive evidence in supermarkets, as manufacturers never rule out explicitly higher types if not forced to do so by law. In the orange example, $m^*(\text{Puglia}) = \text{“Origin: Italy”}$, while the origin of Sicilian oranges is always explicitly stated in Italian supermarkets, namely $m^*(\text{Sicily}) = \text{“Origin: Sicily”}$.

As for consumer misperceptions, the striking fact in (III) is that all processes benefiting from misperceptions are perceived equally. Indeed, if $\hat{\pi}^k(\theta') > \theta'$ then $\hat{\pi}^k(\theta) > \theta$ for $\theta < \theta'$ and so $\hat{\pi}^k(\theta) = \hat{\pi}^k(\theta')$. This creates important issues of “lemons” in the green market, as shown in Section 8. The reason for this is that the unraveling proceeds top-down, so the inferences about the lower types require a higher level of reasoning. Also, misperceptions diminish, but persist until higher levels of reasoning for the dirtier processes. The highest type never benefits from misperceptions; the second-highest type only from level $k < 3$ reasoning; and so on.

I may illustrate Theorem 1 with the case of yogurts. In Italian supermarkets, consumers prefer the local Italian milk, and prefer European milk to milk from more distant locations, for ecological reasons. Naturally, the origin of milk has no effect on the experience of consumers, so that wishful belief formation is predicted. For yogurts, the origin of milk is usually not specified (i.e., $m^*(\theta) = \Theta$), meaning by Theorem 1 that θ is Australia, China, or USA, for instance. The idea of $m^*(\theta)$ is to let the commodity be as “metamorphic” as possible, so that it can metamorphose optimally when facing consumers of every level of reasoning, as illustrates the following diagram:

While it may seem obvious and intuitive that bounded reasoning leads to consumer misperceptions, it is far from necessary. In our case, it results from the combination of two subtle forces: (i) consumers being subject to wishful belief formation under subjective uncertainty; and (ii) the resulting incentives of the seller to use vagueness strategically to maximize the “upward” subjective uncertainty of consumers. If the

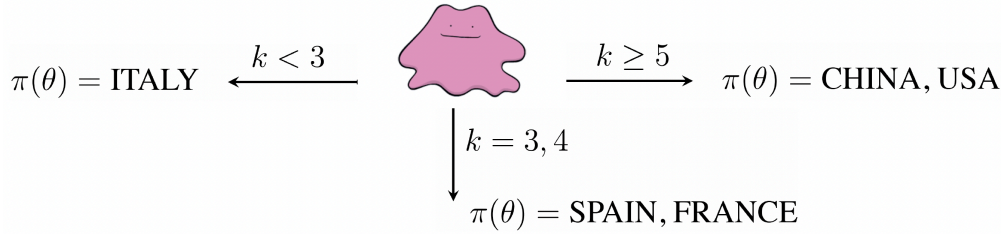


Figure 1. A yogurt made of milk produced outside Europe as perceived in Italy.

seller disclosed $\{\theta\}$ (a rationalizable label), then consumers would not be subject to misperceptions, even under wishful belief formation.

To further see this point, note that under pessimistic belief formation, bounded reasoning does not create misperceptions. From Proposition 1, $\min m \in \Theta^{\Delta, k}(m)$ for all m and k , and so regardless of the degree k of reasoning of consumers, pessimistic belief formation leads to $\bar{v}^k(m) = v(\min m)$ for every m . Thus the uncertainty of the seller is a singleton $\bar{S}_2^{\Delta, *}$ containing the pessimistic strategy, against which a label m is optimal for type θ if and only if $\min m = \theta$.

Proposition 2. *When consumers are subject to pessimistic belief formation, disclosing m is $\bar{\Delta}$ -rationalizable for θ if and only if $\min m = \theta$, and consumers have correct perceptions $\bar{\pi}^k(\theta) = \theta$ for all θ , regardless of the degree k of reasoning in which they engage.*

So, although sellers may disclose the vague labels $m^*(\theta)$ under both belief formation phenomena, under pessimistic belief formation this implies no misperceptions, while under wishful belief formation important issues of overestimation of the value of dirty processes are predicted for consumers engaging in low levels of reasoning.

Finally, another point deserves comment. Milgrom & Roberts (1986) show in an equilibrium analysis (which does not model strategic reasoning explicitly) that the presence of naive consumers does not impede sophisticated consumers to make sharp strategic inferences. In our model of strategic reasoning, the full unraveling obtains as the result of common strong certainty of rationality, an assumption that fails in the presence of naive consumers. If a low level of reasoning is present, sophisticated consumers will be forced to stop their reasoning procedure at a low k , possibly inducing important misperceptions by Theorem 1—in stark contrast with the result of Milgrom & Roberts. On the other hand, sophisticated consumers with a broader vision who understand the strategic incentive of the seller to be robust to all possible models of strategic reasoning of consumers will indeed be able to have correct perceptions since there is a strategic separation of the seller's types, consistently with the spirit of Milgrom & Roberts' result.

This latter conclusion, however, does not hold in all disclosure settings. In practice and in the experimental evidence of Jin, Luca & Martin (2021) and Montero & Seth (2021), sellers often choose to stay silent, including those that are not the lowest type. In our model, the strategic separation of types under uncertain reasoning happens because sellers can partially disclose $m^*(\theta) = [\theta' \geq \theta]$ and optimally mislead this way consumers of all levels of reasoning. In practice, such partial disclosure is not always feasible, and often the seller can only fully disclose or stay silent, i.e.,

$$\tilde{M} = \{\Theta, \{\theta^1\}, \dots, \{\theta^K\}\}.$$

This is a feature of many models of disclosure (e.g., Verrecchia 1983; Dye 1985; Bond & Zeng 2022), and in particular, it is the experimental setting of Jin, Luca & Martin (2021) and Montero & Seth (2021). In this setting the unraveling result obtains as well, and the unique Δ -rationalizable behavior is full disclosure for all types $\theta > \theta^1$. The strategic inferences after observing nondisclosure (let $m^* = \Theta$) are for all $n \geq 0$ as follows:

$$\Theta^{\Delta, 2n+1}(m^*) = \{\theta^1, \dots, \theta^{K-n}\}.$$

But then, the wishful strategy of level- $2n + 1$ is $\hat{v}^{2n+1}(\{\theta\}) = v(\theta)$ for all θ and

$$\hat{v}^{2n+1}(m^*) = v(\max\{\theta^1, \theta^{K-n}\}).$$

Thus, the only way for the seller to take advantage of consumers' bounded reasoning is *nondisclosure*. This means that for all types $\theta < \theta^K$, nondisclosure is $\hat{\Delta}$ -rationalizable and is justified by any belief assigning sufficiently high probability to level-1 reasoning, since $\hat{v}^1(m^*) = v(\theta^K) > v(\theta) = \hat{v}^1(\{\theta\})$. Thus:

Proposition 3. *In the model with feasible disclosures \tilde{M} , nondisclosure is $\hat{\Delta}$ -rationalizable for all sellers of type $\theta < \theta^K$. The unique $\hat{\Delta}$ -rationalizable disclosure of θ^K is full disclosure, $\{\theta^K\}$. Then, upon observing nondisclosure the subjective uncertainty of sophisticated consumers who anticipate the strategic use of vagueness is $\Theta^{\hat{\Delta}, \infty}(m^*) = \{\theta^1, \dots, \theta^{K-1}\}$.*

This shows that the presence of “naive” consumers can create a near-complete failure of unraveling in a model where all the assumptions for full unraveling are satisfied, even for a sophisticated consumers who correctly appraise the full strategic situation of the seller. All types but the highest may find it advantageous to stay silent with the prospect of misleading naive consumers. Notice that for nondisclosure to be justified, the conditions are weaker for the lower types. For type θ^{K-1} it requires the seller to assign sufficiently high probability to level-1 reasoning; for type θ^{K-2} , sufficiently high probability to level $k \leq 3$ reasoning; and so on. So, the lower types are the most likely to find nondisclosure advantageous, and the lower the average level of sophistication, the more types will find it advantageous to stay silent.

The results on overoptimism about hidden information in Jin, Luca & Martin (2021) and Montero & Seth (2021) are thus not necessarily to be viewed as failures of consumers to appraise the strategic incentives of sellers. This may explain why Li & Schipper (2020), whose experimental setting is the one of Theorem 1, find a higher degree of unraveling: in that setting, the presence of naive consumers does not prevent sophisticated consumers with a complete vision of the strategic situation of the seller from making sharp strategic inferences.

7. THE ROLE OF INATTENTIVE CONSUMERS

In the previous section, we have seen that when consumers engage in insufficient strategic reasoning, major issues of misperceptions about the dirty processes arise in the presence of wishful belief formation. Yet, when all consumers engaged in infinite strategic reasoning, consumers did not suffer misperceptions. I shall show in this section that even this latter prediction is highly unlikely. In the presence of inattentive consumers in supermarkets, major misperceptions are predicted *including under infinite strategic reasoning*. To isolate the role of this phenomenon, I shall assume that *all* attentive consumers engage in unbounded reasoning.

The presence of consumers inattentive about certain attributes is highly likely in supermarkets. The attributes related to externalities do not affect material experience, which makes them less salient and hence less likely to come to consumers' mind.³⁰ This means that in the absence of cues recalling their existence, some consumers may not recall to check these attributes and unconsciously assume some "prior" grade.³¹ In contrast, one hardly forgets to check the flavor of a yogurt, even in the absence of cues, since the consumer's experience greatly depends on it.

To formalize these intuitions, suppose that a proportion $\tilde{p} \in (0, 1)$ of consumers is inattentive about attribute θ when not cued by some information (i.e., when $m = \Theta$) and unconsciously perceive type $\tilde{\theta}$ suggested by the framing.³² When m contains some statement on θ , i.e., when $m \subset \Theta$, all consumers are attentive about θ and reason strategically. This captures the idea that the absence of information on an attribute renders that attribute less salient, so that it may not come to consumers' mind when examining the product.

We can incorporate inattentive consumers into the model by modifying the seller's payoff of not disclosing information:

$$(7.1) \quad u_1(\theta, \Theta, \hat{v}) = \tilde{p}f(v(\tilde{\theta})) + (1 - \tilde{p})f(\hat{v}),$$

³⁰The role of salience in shaping attention and recall is abundantly documented; see Bordalo, Gennaioli & Shleifer (2020) and the references therein.

³¹Such assumptions may be formed through unconscious processes influenced by the framing, such as cued recall or stereotypical judgment (see, e.g., Gennaioli & Schleifer 2010; Bordalo, Gennaioli & Shleifer 2013, 2020). This creates scope for framing strategies to shape consumers' beliefs.

³²The results of this section obtain also if inattentive consumers perceive a distribution $\tilde{\mu} \in \Delta(\Theta)$, by reinterpreting $v(\tilde{\theta})$ in (7.1) as $v(\tilde{\theta}) = \mathbb{E}_{\tilde{\mu}}[v(\cdot)]$.

that is, the seller's payoff of being silent on the attribute depends also on the value perceived by inattentive consumers. I shall also assume that $f(\cdot)$ is *continuous*.

The analysis concerns the perceptions of consumers attentive about θ and the seller's disclosure behavior in this environment with inattentive consumers.

The key informational issue that arises is that, while all types $\theta \geq \tilde{\theta}$ are unraveled by strategic reasoning, all types θ worse than $\tilde{\theta}$ can rationalizably find silence advantageous: Whenever the seller believes that attentive consumers assign sufficiently high probability to $\max\{\theta : \theta < \tilde{\theta}\}$ after observing nondisclosure, the extra utility of the inattentive consumers who perceive $v(\tilde{\theta}) > v(\theta)$ more than compensates the penalization of mild skepticism when $\mu^2(\theta^1|\Theta)$ is sufficiently small.

Proposition 4. *For all sellers of type $\theta < \tilde{\theta}$, nondisclosure is Δ -rationalizable, while all sellers of types $\theta \geq \tilde{\theta}$ disclose m such that $\min m = \theta$.*

This means that upon observing nondisclosure, sophisticated consumers are unable to make sharp inferences about the seller's type because nondisclosure is consistent with strategic behavior of every $\theta < \tilde{\theta}$, i.e., $\Theta^{\Delta, \infty}(\Theta) = [\theta^1, \tilde{\theta}]$. Then, sophisticated consumers will perceive the highest type just below the prior $\tilde{\theta}$ of inattentive consumers. Understanding this, all sellers manufacturing types that are worse than $\tilde{\theta}$ will find nondisclosure advantageous. (I assume the seller best responds to the sophisticated wishful strategy.)

Theorem 2. *Under wishful belief formation, the following happens:*

- (I) *For all sellers' types $\theta < \tilde{\theta}$, nondisclosure is the unique $\hat{\Delta}$ -rationalizable behavior, i.e., $m^*(\theta) = \Theta$. For the types $\theta \geq \tilde{\theta}$, they disclose m such that $\min m = \theta$.*
- (II) *For every process $\theta < \tilde{\theta}$, sophisticated consumers perceive $\hat{\pi}^\infty(\theta) = \max\{\theta' : \theta' < \tilde{\theta}\}$ and inattentive consumers perceive $\tilde{\pi}(\theta) = \tilde{\theta}$. For the other processes, sophisticated consumers have correct perceptions and there are no inattentive consumers.*

This gives considerable scope for the dirtiest producers to be perceived cleaner than they are, including by sophisticated consumers. Even an arbitrarily small proportion of inattentive consumers about attribute θ may generate major misperceptions for the attentive ones.³³

This mechanism explains why we are unable to infer the country of cultivation of the organic honey of Section 2. In this example $\tilde{\theta}$ is the highest type, France, so that nondisclosure is consistent with strategic behavior for all other possible types;

³³Again, wishful belief formation is key for this prediction. It can be shown that for sufficiently small \tilde{p} , under pessimistic belief formation all sellers have incentive to distinguish themselves from the lower types, and attentive consumers have correct perceptions.

so consumers engaging in sophisticated reasoning believe the second-highest type of the seller due to wishful belief formation.

In fact, it is precisely those goods that have *some* green attribute that will be most able to (and hence will) mislead consumers about the other attributes. Putting forward the green attribute suggests a high $\tilde{\theta}$ for the other attributes and diminishes the proportion of attentive consumers. For example, the organic label suggests consumers that “everything is green” and inhibits critical thinking about other aspects of production processes; organic products are often *a priori* perceived local, while the truth is often surprisingly different.

Note that the payoff of sellers choosing nondisclosure (the dirtiest types) increases in the proportion of inattentive consumers and in $\tilde{\theta}$. Their marketing departments will thus employ their genius to design labels maximizing inattention about θ and creating positive framing effects.

8. MISPERCEPTIONS AND UNRAVELING OF THE GREEN MARKET

The results of the previous sections identify optional disclosure as a major source of environmental market failure. If the level of strategic reasoning in which consumers engage is low or if there are inattentive consumers, we can expect an important unraveling of the “green” market. To visualize this point, I introduce in this section a simple model of consumer choice where perceptions about the alternatives are endogenous and determined by the disclosure interaction. (In particular, I depart from the classical “revealed preferences” assumption according to which consumer choice reflects true preferences.)

Consider a supermarket with a set I of consumers and a commodity, say honey, that can be produced with processes $\theta^1 < \theta^2 < \theta^3$. The order reflects environmental concern associated with the production processes (the lowest type is dirty and the highest one is green). Specifically, the preferences of consumer i are described by $v_i(\theta^1) < v_i(\theta^2) < v_i(\theta^3)$. Assume that prices $p(\theta)$ are competitive and set at marginal cost $c(\theta)$. Assuming that costs are greater for the green alternatives, we have $p(\theta^1) < p(\theta^2) < p(\theta^3)$. The alternatives are strategically labeled and the consequent perceptions of consumer i are $\hat{\pi}^i(\theta)$ for each θ . As obtains in many standard models of consumer choice for goods that are substitutes (here different alternatives of honey), consumer i buys the alternative with the highest *perceived* value-price ratio, namely:

$$(8.1) \quad C_i(v_i, \hat{\pi}^i, p) = \arg \max_{\theta} \frac{v_i(\hat{\pi}^i(\theta))}{p(\theta)}.$$

If $\hat{\pi}^i(\theta) = \theta$ for all θ , we obtain the standard model of consumer choice where choices reflect true preferences. It is convenient to denote the choices that consumers would make under correct perceptions as $C_i^*(v_i, p)$ (e.g., under mandatory disclosure of processes).

Then the demand of alternative θ under voluntary disclosure is given by

$$D(\theta) = \{i \in I : C_i(v_i, \hat{\pi}^i, p) = \theta\},$$

and I denote by $D^*(\theta) = \{i \in I : C_i^*(v_i, p) = \theta\}$ the demand of θ under correct perceptions.

The results on consumer misperceptions imply major issues of “lemons” in the green market. First, from Theorem 1 and Theorem 2, all consumers i that are inattentive and framed to think $\hat{\theta} = \theta^3$, or engage in level $k < 3$ reasoning will perceive the dirty good as ecological, i.e., $\hat{\pi}^i(\theta^1) = \theta^3$, and will therefore buy the dirty good since cheaper, even those for which $C_i^* > \theta^1$, absorbing the demand of θ^2 and θ^3 . If the proportion of such consumers is high, this may considerably reduce the size of the green market. As an extreme case, if all consumers are inattentive or engage in insufficient reasoning, we have a complete unraveling of the green market, with $D(\theta^2) = D(\theta^3) = \emptyset$, even if consumers have highly ecological preferences and $D^*(\theta^1) = \emptyset$. This unraveling effect arises from an informational issue akin to the “lemons” problem of Akerlof (1970): the commodities are all perceived equally by consumers, and so only the cheapest is selected.

An even more subtle unraveling effect happens with consumers who are attentive and engage in infinite reasoning. Since inattentive consumers perceive $\hat{\theta} = \theta^3$, by Theorem 2 sophisticated consumers will perceive $\hat{\pi}^i(\theta^1) = \hat{\pi}^i(\theta^2) = \theta^2$. This has two effects. First, process θ^2 is driven out of the market since θ^1 is cheaper and equally perceived, so that $D(\theta^2) = \emptyset$. The true demand of θ^2 is entirely absorbed by θ^1 , that is, $D^*(\theta^2) \subseteq D(\theta^1)$. Second, since θ^1 is perceived as θ^2 and θ^1 is cheaper than θ^2 , the “true” incentive compatibility condition of θ^2 with θ^3 is relaxed. This implies that part of the true demand $D^*(\theta^3)$ will shift to $D(\theta^1)$. To see this, although all $i \in D^*(\theta^3)$ satisfy $v_i(\theta^3)/p(\theta^3) > v_i(\theta^2)/p(\theta^2)$, we may have $v_i(\theta^3)/p(\theta^3) < v_i(\theta^2)/p(\theta^1)$ so that $i \in D(\theta^1)$. This effect will be particularly strong if θ^1 is much cheaper than θ^2 .

This last unraveling effect captures the essence of the issue raised by misperceptions on dirty processes under voluntary disclosure. The mechanism differs from the traditional “lemons” in that the dirty and the green alternatives are *differentiated* since $\hat{\pi}^i(\theta^1) < \hat{\pi}^i(\theta^3)$. Yet, the “dirty” process benefits from misperceptions that makes it more likely to beat the value-price ratio of the ecological alternatives. This is another form the “unraveling” mechanism of Akerlof (1970) can take, where high and low quality producers are differentiated, but high-quality producers lose market shares because of a *reduced perceived difference* with the low-quality producers.³⁴

Another difference with the “lemons” market of Akerlof is that, due to the absence of feedback on processes, the misperceptions persist even after most green processes have been driven out of the market. This means that the “dirty” goods will continue

³⁴Akerlof’s “lemons” problem is a special case where consumers cannot perceive any difference between high and low quality sellers.

to absorb the demand of “green” producers. As a consequence, green processes will be persistently less rewarded by the market, which depresses incentives to invest in green R&D and green technology adoption.

9. CONCLUSIONS

I have analyzed the perceptions consumers form about production processes given the information disclosed by manufacturers. Although the clean processes are signaled in markets, little information about the dirty processes is transmitted. Positive and experimental evidence suggest that in supermarkets, inattention to production processes and failure of consumers to appraise the strategic disclosure incentives of manufacturers is the norm, rather than the exception. Due to the psychology of wishful belief formation, there is then ground to think that underestimation of environmental externalities is pervasive and substantial.

The fact that green processes are signaled in supermarkets does not suffice for environmental preferences to be reflected in consumer choice. It is well documented in the literature on the psychology of value that individuals are more sensible to negative information than to positive information due to the phenomenon of *loss aversion* (Kahneman & Tversky 1984). Thus, the problem of misperceptions on “dirty” goods particularly favor their election relative to the informed situation, and we can expect mandatory disclosure to generate a strong negative demand shock for dirty goods. This expectation is strongly supported by experimental and empirical evidence of Kim, Kim & Arora (2022) and Mathios (2000).

These results have important descriptive and policy implications. Under voluntary disclosure, marketing tricks become a cheap substitute to green innovation and technology adoption. The sellers being able to be perceived green on certain attributes without being so, being *actually* green on such attributes is not economically viable for producers. The mechanism of Aghion et al. (2022) can thus be severely impaired under voluntary disclosure, with an environmental market failure even when consumers have strong environmental concerns. Shedding light on processes will drastically change the incentives of producers—the ability to produce *cleanly at a low price* will become the key competitive advantage on the market. This will result in a competition on processes incentivizing firms to develop clean production technologies that are less costly in order to attract shares of markets.³⁵

This represents a mild and virtually costless intervention relying on decentralized mechanisms. As for its practical implementation, the creation of a mandatory *universal label* where some key properties of processes are stated unambiguously would be efficacious. A standardized label would facilitate information acquisition and the comparison of alternatives on the basis of processes, and avoid the manufacturers’ strategic use of complex disclosure (Jin, Luca & Martin 2022). Regulators

³⁵This will make green alternatives affordable for a larger share of the population—a necessity for the ecological transition to occur.

should target disclosure of aspects of production that are simple to verify and importantly connected to externalities (even if indirectly), such as the origin of production of ingredients. Such regulations will also empower more traditional environmental policies aimed at incentivizing firms to adopt and develop clean technologies by increasing firms' intrinsic market stake in their use.

Despite the need of acting quickly to limit the now unavoidable climate change, political constraints are important and greatly slow down the implementation of decisive public interventions. In a situation where nations fail to reach an ambitious climate agreement and where carbon taxation faces harsh opposition by citizens, it is essential to develop alternative, high-impact interventions toward the green transition that could be implemented rapidly by any nation. Based on the results of Aghion et al. (2022) and the present analysis, restoring the pressure of the demand side on dirty production by regulating information disclosure about production processes in supermarkets seems a promising measure in this sense.

APPENDIX: PROOFS

Preliminaries. I shall introduce some convenient notation. Given a measurable space (X, \mathcal{S}_X) and an event $Q \in \mathcal{S}_X$, I let $\Delta(Q) := \{\mu \in \Delta(X) : \mu(Q) = 1\}$. Let $\Sigma_1^{\Delta, k}(\theta) := \{m \in M : (\theta, m) \in \Sigma_1^{\Delta, k}\}$ denote the set of (Δ, k) -rationalizable labels for type θ . Further, let $M(\theta) := \{m \in M : \theta \in m\}$ denote the set of truthful labels for type θ . Finally, let $U_1 : \Theta \times M \times \Delta(S_2) \rightarrow \mathbb{R}$ denote the expected payoff function of the seller, which, for beliefs with finite support and truthful labels $m \in M(\theta)$ is:

$$U_1(\theta, m, \mu^1) := \sum_{\hat{v} \in \text{Supp } \mu^1} f(\hat{v}(m)) \mu^1(\hat{v}).$$

PROOF OF PROPOSITION 1:

I shall show a bit more than Proposition 1 and characterize the whole sequence of (Δ, k) -rationalizable labels, which will be essential to prove Theorem 1.

Step 1. For the *seller*, due to the disclosure laws on truthful reporting only truthful labels are justifiable, and every truthful label is justifiable, since $BR_1(\theta, \mu^1) = M(\theta)$ for any μ^1 assigning probability one to a constant valuation strategy. So, $\Sigma_1^{\Delta, 1}(\theta) = M(\theta)$ for each θ .

As for the *consumer*, for our purposes it is enough to note that $S_2^{\Delta, 1}$ contains constant strategies, as there are beliefs in Δ_2 justifying constant valuation strategies (e.g., μ^2 with $\mu^2(\theta|m) = 1/K$ for all θ and m satisfies mild skepticism).

Step 2. For the *consumer*, note that for every m we have $\Theta^{\Delta, 1}(m) = m$. Thus, for each $\hat{v} \in S_2^{\Delta, 2}$ there exists a belief system $\mu^2 \in \Delta_2$ such that $\mu^2(m|m) = 1$ for all m , for which $\hat{v} \in BR_2^*(\mu^2)$. The essential implications are that: (a) $\hat{v}(\{\theta\}) = v(\theta)$ for each θ ; and (b) for all $m \in M(\theta)$ such that $\max m = \theta$, and $\min m < \theta$, we have

$\hat{v}(m) < v(\theta)$ since:

$$\begin{aligned}\hat{v}(m) &= BR_2(\mu^2(\cdot|m)) = \sum_{\theta' \in m} v(\theta') \mu^2(\theta'|m) \\ &\leq v(\min m) \mu^2(\min m|m) + v(\theta)(1 - \mu^2(\min m|m)) \\ &< v(\theta),\end{aligned}$$

where the first inequality is because $\max \text{Supp } \mu^2(\cdot|m) \leq \theta$ since $\max m = \theta$ and $\mu^2(m|m) = 1$; the second inequality is because $\min m < \theta$ and $\mu^2(\min m|m) > 0$ by mild skepticism.

For the *seller*, since $S_2^{\Delta,1}$ contains constant strategies, it is immediate that $\Sigma_1^{\Delta,2}(\theta) = \Sigma_1^{\Delta,1}(\theta) = M(\theta)$ for each θ .

Since $\Sigma_1^{\Delta,2} = \Sigma_1^{\Delta,1}$, we have $S_2^{\Delta,k+1} = S_2^{\Delta,k}$ for $k > 2$ even and $\Sigma_1^{\Delta,k+1} = \Sigma_1^{\Delta,k}$ for $k > 2$ odd. We conclude the proof with an inductive argument on $n \in \mathbb{N}$:

Inductive hypothesis (n): If $\theta > \theta^{K-n}$, $\Sigma_1^{\Delta,2n+1}(\theta) = \{m : \min m = \theta\}$, and $\Sigma_1^{\Delta,2n+1}(\theta) = M(\theta) \setminus \{m : \min m < \theta \wedge \max \Theta^{\Delta,2n-1}(m) = \theta\}$ otherwise.

Basis step ($n = 1$). Fix $\theta \in \Theta$, $\mu^1 \in \Delta(S_2^{\Delta,2})$, and $m \in \{m : \min m < \theta, \max m = \theta\}$. We show that $m \notin BR_1(\theta, \mu^1)$. Notice that μ^1 assigns probability one to strategies \hat{v} that satisfy (a) $\hat{v}(\{\theta\}) = v(\theta)$; and (b) $\hat{v}(m) < v(\theta)$. But then $\{\theta\}$ yields a higher expected payoff than m for type θ under μ^1 , and thus $m \notin BR_1(\theta, \mu^1)$. Since $\mu^1 \in \Delta(S_2^{\Delta,2})$ was arbitrary, it follows that $m \notin \Sigma_1^{\Delta,3}(\theta)$. For $\theta = \theta^K$, this implies that $\Sigma_1^{\Delta,3}(\theta^K) = \{\{\theta^K\}\} = \{m : \min m = \theta^K\}$. Further, for all $\theta \neq \theta^K$ it can be shown that every $m' \in M(\theta) \setminus \{m : \min m < \theta \wedge \max m = \theta\}$ can be justified as a best reply to some $\mu^1 \in \Delta(S_2^{\Delta,2})$.

Inductive step ($n + 1$). Fix $\theta \in \Theta$, $\mu^1 \in \Delta(S_2^{\Delta,2(n+1)})$, and $m \in \{m' : \min m' < \theta, \max \Theta^{\Delta,2n+1}(m') = \theta\}$. We show that $m \notin BR_1(\theta, \mu^1)$. As before, for all $\hat{v} \in S_2^{\Delta,2(n+1)}$ there is $\mu^2 \in \Delta_2$ such that $\mu^2(\Theta^{\Delta,2n+1}(m')|m')$ for every m' and $\hat{v} \in BR_2^*(\mu^2)$. Since $\max \Theta^{\Delta,2n+1}(m) = \theta$, $\text{Supp } \mu^2(\cdot|m) \subseteq \{\theta^1, \dots, \theta\}$ and since $\min m < \theta$ and $\min m \in \Theta^{\Delta,2n+1}(m)$ by the inductive hypothesis, by mild skepticism we have $\hat{v}(m) < v(\theta) = \hat{v}(\{\theta\})$. So $m \notin BR_1(\theta, \mu^1)$. Therefore, for all θ we have

$$\Sigma_1^{\Delta,2(n+1)+1}(\theta) \subseteq M(\theta) \setminus \{m : \min m < \theta \wedge \max \Theta^{\Delta,2n+1}(m) = \theta\}.$$

We show the other inclusion: Fix $m \in M(\theta) \setminus \{m : \min m < \theta \wedge \max \Theta^{\Delta,2n+1}(m) = \theta\}$. We show that $m \in \Sigma_1^{\Delta,2(n+1)+1}$. If $\min m = \theta$, then m is a best reply against the skeptical strategy $\hat{v} \in S_2^{\Delta,2(n+1)}$ with $\hat{v}(m') = v(\min m')$ for all $m' \in M$. If $\min m < \theta$ and $\theta^* = \max \Theta^{\Delta,2n+1}(m) > \theta$, then m is a best reply to the strategy $\hat{v} \in S_2^{\Delta,2(n+1)}$ with $\hat{v}(m) = \varepsilon v(\min m) + (1 - \varepsilon)v(\theta^*)$ and $\hat{v}(m') = v(\min m')$ for $m' \neq m$. As $\varepsilon \rightarrow 0$ we have $\hat{v}(m) \rightarrow \hat{v}(\theta^*) > v(\theta) \geq \hat{v}(m')$ for all $m' \in M(\theta) \setminus \{m\}$, justifying m as a best reply for $\varepsilon > 0$ sufficiently small.

Finally, note that if $\theta > \theta^{K-(n+1)}$, then for all $m \in M(\theta)$ we have $\max \Theta^{\Delta, 2n+1}(m) \leq \theta$. Indeed, by the inductive hypothesis for all $\theta' > \theta \geq \theta^{K-n}$ we have $\Sigma_1^{\Delta, 2n+1}(\theta') = \{m' : \min m' = \theta'\}$. Since $\theta \in m$, we have $\min m \leq \theta < \theta'$ and so $m \notin \Sigma_1^{\Delta, 2n+1}(\theta')$. Hence, $\theta' \notin \Theta^{\Delta, 2n+1}(m)$. It follows that any $m \in M(\theta)$ such that $\min m = \theta$ yields $\hat{v}(\theta) = v(\theta)$ and so $m \in BR_1(\theta, \mu^1)$ given the previous point. So, $\Sigma_1^{\Delta, 2(n+1)+1}(\theta) = \{m : \min m = \theta\}$. ■

PROOF OF THEOREM 1:

From the proof of Proposition 1, we have the following characterization of the (Δ, k) -rationalizable labels m as a function of types and reasoning. First, $\Theta^{\Delta, 1}(m) = \Theta^{\Delta, 2}(m) = m$. Second, for all $n \geq 1$ and $\theta > \theta^{K-n}$ we have

$$(9.1) \quad \Sigma_1^{\Delta, 2n+1}(\theta) = \{m \in M : \min m = \theta\},$$

which has an implication we shall note for future reference:

$$(9.2) \quad \forall m \in \{m' : \min m \leq \theta^{K-n}\}, \quad \Theta^{\Delta, 2n+1}(m) \subseteq m \cap \{\theta^1, \dots, \theta^{K-n}\}.$$

Third, for all $n \geq 1$, for all $\theta \leq \theta^{K-n}$ we have

$$(9.3) \quad \Sigma_1^{\Delta, 2n+1}(\theta) = M(\theta) \setminus \{m : \min m < \theta \wedge \max \Theta^{\Delta, 2(n-1)+1}(m) = \theta\}.$$

With this, for our purposes it is enough to show the following characterization of the strategic inferences upon observing the labels $m^*(\theta)$.

Claim 1: for all θ and $n \geq 0$, $\Theta^{\Delta, 2n+1}(m^*(\theta)) = \{\theta\} \cup [\theta, \theta^{K-n}]$.

Proof: We prove the claim by induction. The basis for $n = 0$ is immediate since $\Theta^{\Delta, 1}(m^*(\theta)) = m^*(\theta) = [\theta, \theta^K]$. Suppose by way of induction that Claim 1 holds for $n \geq 0$; we show that it holds for $n + 1$. First, since $\min m^*(\theta) = \theta$ we have $m^*(\theta) \in M(\theta)$ and from (9.3) we have $m^*(\theta) \in \Sigma_1^{\Delta, 2(n+1)+1}(\theta)$ and so $\theta \in \Theta^{\Delta, 2(n+1)+1}(m^*(\theta))$. For $\theta' > \theta$, we have (i) $\theta' \in m^*(\theta)$, so that $m^*(\theta) \in M(\theta')$; and (ii) $\min m^*(\theta) < \theta'$. From (i) and (ii), it follows from (9.3) that $m^*(\theta) \in \Sigma_1^{\Delta, 2(n+1)+1}(\theta')$ if and only if $\theta' < \max \Theta^{\Delta, 2n+1}(m^*(\theta)) = \theta^{K-n}$, where the last equality holds by the inductive hypothesis. Thus, $\Theta^{\Delta, 2(n+1)+1}(m^*(\theta)) = \{\theta\} \cup [\theta, \theta^{K-(n+1)}]$. □

We show part (I), namely that label $m^*(\theta)$ is weakly dominant for θ with respect to $S_2^{\Delta, *}$. To ease notation, let $m^* = m^*(\theta)$. First, notice that $m^* \in M(\theta)$ does weakly better than all other disclosures against every $\hat{v}^k \in S_2^{\Delta, *}$, since:

$$\forall m \in M(\theta), \quad \hat{v}^k(m^*) = v(\max \Theta^{\Delta, k}(m^*)) \geq v(\max \Theta^{\Delta, k}(m)) = \hat{v}^k(m),$$

where the inequality holds because from Claim 1 and (9.2) we have $\max \Theta_1^{\Delta,k}(m^*) \geq \max \Theta_1^{\Delta,k}(m)$. Now, consider any $m \neq m^*$. If there is $\theta' \in m$ such that $\theta' < \theta$, then for $k = \infty$ we have

$$\max \Theta^{\Delta,\infty}(m^*) = \theta > \theta' \geq \max\{\min m\} = \max \Theta^{\Delta,\infty}(m),$$

and thus m yields a strictly lower payoff than m^* against $\hat{v}^\infty \in S_2^{\Delta,*}$. If there exists $\theta' \notin m$ such that $\theta' > \theta$, take k such that $\max \Theta^{\Delta,k}(m^*) = \theta'$, which must exist by Claim 1. We then have $\max \Theta^{\Delta,k}(m^*) = \theta' > \max \Theta^{\Delta,k}(m)$, so that m yields a strictly lower payoff than m^* against $\hat{v}^k \in S_2^{\Delta,*}$.

Finally, (II) follows from Claim 1; (III) follows from (II) and wishful belief formation. ■

PROOF OF PROPOSITION 4:

The proof of the unraveling for the types $\theta \geq \tilde{\theta}$ is analogous to the proof of Proposition 1. The only point that deserves comment is that once the unraveling of the types $\theta > \tilde{\theta}$ has been incorporated in consumers' conditional beliefs, type $\tilde{\theta}$ has a strict incentive to disclose m with $\min m = \tilde{\theta}$ instead of $\tilde{m} = \Theta$ because, compared to m , attentive consumers penalize \tilde{m} due to mild skepticism (by assigning positive probability to θ^1).

Suppose we are at step N , where the unraveling of the types $\theta \geq \tilde{\theta}$ is already incorporated into the consumers' conditional beliefs, i.e., $\Theta^{\Delta,N-1}(m) \subseteq \{\theta : \theta < \tilde{\theta}\}$ for all m with $\min m < \tilde{\theta}$. Let $\bar{\theta} := \max\{\theta : \theta < \tilde{\theta}\}$. For any type $\theta < \tilde{\theta}$, the maximal payoff that can be achieved when disclosing a truthful $m' \subset \Theta$ is $\pi \leq f(v(\bar{\theta}))$, because $\hat{v}(m') \leq v(\bar{\theta})$ for every $\hat{v} \in S_2^{\Delta,N}$. On the other hand, by disclosing $m = \Theta$, since $\bar{\theta} \in \Theta^{\Delta,N-1}(m)$, the consumer may hold a conjecture μ^2 such that $\mu^2(\bar{\theta}|m) = 1 - \varepsilon$ and $\mu^2(\theta^1|m) = \varepsilon$, so that $\mu^2 \in \Delta_2$. Then, there is $\hat{v} \in S_2^{\Delta,N}$ with

$$\hat{v}(m) = (1 - \varepsilon)v(\bar{\theta}) + \varepsilon v(\theta^1).$$

Take $\mu^1 \in \Delta(S_2^{\Delta,N})$ such that $\mu^1(\hat{v}) = 1$. Let $\kappa \equiv f(v(\tilde{\theta})) - f(v(\bar{\theta})) > 0$. For $\varepsilon \rightarrow 0$ we have $\hat{v}(m) \rightarrow v(\bar{\theta}) \geq \hat{v}(m')$, and so by continuity of $f(\cdot)$, for all $\varepsilon > 0$, there is $\varepsilon > 0$ sufficiently small such that $f(\hat{v}(m')) - f(\hat{v}(m)) < \varepsilon$. Thus, for ε sufficiently small we have

$$\tilde{p}f(v(\tilde{\theta})) + (1 - \tilde{p})f(\hat{v}(m)) > \tilde{p}f(\hat{v}(m')) + (1 - \tilde{p})f(\hat{v}(m')) = f(\hat{v}(m')),$$

where the inequality holds because for ε sufficiently small we have

$$\tilde{p}\kappa - (1 - \tilde{p})[f(\hat{v}(m')) - f(\hat{v}(m))] > \tilde{p}\kappa - (1 - \tilde{p})\varepsilon > 0.$$

Hence, $m \in BR_1(\theta, \mu^1)$ and $(\theta, m) \in \Sigma_1^{\Delta,N+1}$. The same argument can be repeated in subsequent steps for every $\theta < \tilde{\theta}$, so that $(\theta, m) \in \Sigma_1^{\Delta,\infty}$ for all $\theta < \tilde{\theta}$. ■

PROOF OF THEOREM 2:

I show that (I) the unique best response to the sophisticated wishful strategy is $m^* = \Theta$ for all types $\theta < \tilde{\theta}$, while the other types $\theta \geq \tilde{\theta}$ display m such that $\min m = \theta$. (II) follows from (I). Proposition 4 and the unraveling of the types $\theta \geq \tilde{\theta}$ imply that the sophisticated wishful strategy \hat{v}^∞ satisfies $\hat{v}^\infty(m) = v(\min m)$ for every m such that $\min m \geq \tilde{\theta}$ and $\hat{v}^\infty(m) \leq v(\tilde{\theta})$ for every m such that $\min m < \tilde{\theta}$. Further, since $\bar{\theta} := \max\{\theta : \theta < \tilde{\theta}\} \in \Theta^{\Delta, \infty}(m^*)$, wishful belief formation implies $\hat{v}^\infty(m^*) = v(\bar{\theta})$. But then for every $\theta < \tilde{\theta}$ and $\mu^1 \in \Delta(\{\hat{v}^\infty\})$, for every $m \in M(\theta)$ (necessarily such that $\min m < \tilde{\theta}$) different than m^* we have

$$U_1(\theta, m^*, \mu^1) = \tilde{p}f(v(\tilde{\theta})) + (1 - \tilde{p})f(v(\bar{\theta})) > f(\hat{v}^\infty(m)) = U_1(\theta, m, \mu^1),$$

where the inequality holds because $f(v(\tilde{\theta})) > f(v(\bar{\theta})) \geq f(\hat{v}^\infty(m))$. Therefore, $BR_1(\theta, \mu^1) = \{m^*\}$. Hence, the unique $\hat{\Delta}$ -rationalizable label for every θ such that $\theta < \tilde{\theta}$ is $m^* = \Theta$. As for $\theta \geq \tilde{\theta}$, we have by Proposition 4 that if $\min \tilde{m} < \theta$ then $\Theta^{\Delta, \infty}(\tilde{m}) \subseteq \{\theta^1, \dots, \bar{\theta}\}$, while for every m such that $\min m = \theta$ we have $\Theta^{\Delta, \infty}(m) = \{\theta\}$. Thus $\hat{v}^\infty(m) = v(\theta) > v(\bar{\theta}) \geq \hat{v}^\infty(\tilde{m})$. So for $\mu^1 \in \Delta(\{\hat{v}^\infty\})$ we have $BR_1(\theta, \mu^1) = \{m : \min m = \theta\}$. ■

References

- [1] Aghion, P., Bénabou, R., Martin, R. and Roulet, A. 2022. “Environmental Preferences and Technological Choices: Is Market Competition Clean or Dirty?” *American Economic Review: Insights* (forthcoming).
- [2] Akerlof, G. 1970. “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” *Quarterly Journal of Economics* 84 (3): 488-500.
- [3] Akerlof, G. and Dickens, W. 1982. “The Economic Consequences of Cognitive Dissonance.” *American Economic Review* 72: 307-319.
- [4] Alaoui, L. and Penta, A. 2016. “Endogenous Depth of Reasoning.” *Review of Economic Studies* 83: 1297-1333.
- [5] Andre, P., Boneva, T., Chopra, F., Falk, A. 2021. “Fighting Climate Change: The Role of Norms, Preferences, and Moral Values.” mimeo.
- [6] Battigalli, P. 1996. “Strategic Rationality Orderings and the Best Rationalization Principle.” *Games and Economic Behavior* 13: 178-200.
- [7] Battigalli, P. 1997. “On Rationalizability in Extensive Games.” *Journal of Economic Theory* 74 (1): 40-61.
- [8] Battigalli, P. 2003. “Rationalizability in Infinite, Dynamic Games with Incomplete Information.” *Research in Economics* 57 (1): 1-38.
- [9] Battigalli, P. 2006. “Rationalization in Signaling Games: Theory and Applications.” *International Game Theory Review* 8 (1): 67-93.
- [10] Battigalli, P. and Prestipino, A. 2013. “Transparent Restrictions on Beliefs and Forward-Induction Reasoning in Games with Asymmetric Information.” *The B E Journal of Theoretical Economics*, 13 (1): 1-53.
- [11] Battigalli, P. and Siniscalchi, M. 2002. “Strong Belief and Forward Induction Reasoning.” *Journal of Economic Theory* 106: 356-391.
- [12] Battigalli, P. and Siniscalchi, M. 2003. “Rationalization and Incomplete Information.” *Theoretical Economics (Advances)* 1: Article 3.

- [13] Battigalli, P. and Siniscalchi, M. 2007. "Interactive Epistemology in Games with Payoff Uncertainty." *Research in Economics* 61 (4): 165-184.
- [14] Bénabou, R. and Tirole, J. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *The Journal of Economic Perspectives* 30 (3): 141-164.
- [15] Bond, P. and Zeng, Y. 2022. "Silence is Safest: Information Disclosure When the Audience's Preferences are Uncertain." *Journal of Financial Economics* 145: 178-193.
- [16] Bordalo, P., Conlon, J., Kwon, S. Y., Gennaioli, N. and Shleifer, A. 2022. "Memory and Probability." Working Paper.
- [17] Bordalo, P., Gennaioli, N. and Shleifer, A. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121: 803-843.
- [18] Bordalo, P., Gennaioli, N. and Shleifer, A. 2020. "Memory, Attention, and Choice." *Quarterly Journal of Economics* 121: 1399-1442.
- [19] Drobner, C. 2022. "Motivated Beliefs and Anticipation of Uncertainty Resolution." *American Economic Review: Insights* 4: 89-105.
- [20] Dye, R. 1985. "Disclosure of Nonproprietary Information." *Journal of Accounting Research* 23: 123-145.
- [21] Farrell, J. 1986. "Voluntary Disclosure: Robustness of the Unraveling Result, and Comments on its Importance." In *Antitrust and Regulation*, edited by New York, NY: Lexington Books.
- [22] Gennaioli, N. and Shleifer, A. 2010. "What Comes to Mind." *Quarterly Journal of Economics* 125: 1399-1433.
- [23] Grossman, S. 1981. "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24(2): 461-483.
- [24] Grossman, S. and Hart, O. 1980. "Disclosure Laws and Takeover Bids." *Journal of Finance* 25(2): 323-334.
- [25] Hagenbach, J. and Perez-Richet, E. 2018. "Communication with Evidence in the Lab." *Games and Economic Behavior* 112: 139-165.
- [26] Heifetz, A., Meier, M. and Schipper, B. C. 2020. "Prudent Rationalizability in Generalized Extensive-Form Games with Unawareness." *B. E. Journal of Theoretical Economics* 21: 525-556.
- [27] Hill, B. 2021. "Being Up Front about Income Inequality." HEC Paris Research Paper.
- [28] Hill, B. and Lloyd, T. 2021. "Are People Willing To Pay for Reduced Inequalities?" HEC Paris Research Paper.
- [29] Hirshleifer, D., Lim, S. S. and Teoh, S. H. 2004. "Disclosure to an Audience with Limited Attention." mimeo
- [30] Jin, G. Z., Luca, M. and Martin, D. 2021. "Is No News (Perceived As) Bad News? An Experimental Investigation of Information Disclosure." *American Economic Journal: Microeconomics* 13: 141-173.
- [31] Jin, G. Z., Luca, M. and Martin, D. 2022. "Complex Disclosure." *Management Science* 68: 3236-3261.
- [32] Jung, W.-O. and Kwon, Y. 1988. "Disclosure When the Market Is Unsure of Information Endowment of Managers." *Journal of Accounting Research* 26: 146-153.
- [33] Jovanovic, B. 1982. "Truthful Disclosure of Information." *The Bell Journal of Economics* 13: 36-44.
- [34] Kahan, D. M. 2013. "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making* 8: 407-424.
- [35] Kahan, D. M., Dawson, E. C., Peters, E. and Slovic, P. 2017. "Motivated Numeracy and Enlightened Self-Government." *Behavioural Public Policy* 1: 54-86.

- [36] Kahneman, D. and Tversky, A. 1984. "Choices, Values, and Frames." *American Psychologist* 39 (4): 341-350.
- [37] Kim, Y., Kim, S. and Arora, N. 2022. "GMO Labeling Policy and Consumer Choice." *Journal of Marketing* 86: 21-39.
- [38] Koessler, F. and Renault, R. 2012. "When Does a Firm Disclose Product Information?" *The RAND Journal of Economics* 43: 630-649.
- [39] Li, Y. X. and Schipper, B. C. 2018. "Disclosure Under Unawareness: An Experiment." mimeo.
- [40] Li, Y. X. and Schipper, B. C. 2020. "Strategic Reasoning in Persuasion Games: An Experiment." *Games and Economic Behavior* 121: 329-367.
- [41] Manili, J. 2022. "Environmental Preferences and Technological Choices: A Comment on the Role of Consumer Information." mimeo.
- [42] Mathios, A. 2000. "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market." *Journal of Law and Economics* 43: 651-678.
- [43] Milgrom, P. 1981. "Good News and Bad News: Representation Theorems and Applications." *The Bell Journal of Economics* 12: 380-391.
- [44] Milgrom, P. and Roberts, J. 1986. "Relying on the Information of Interested Parties." *The RAND Journal of Economics* 17: 18-32.
- [45] Montero, M. and Sheth, J. 2021. "Naivety about Hidden Information: An Experimental Investigation." *Journal of Economic Behavior and Organization* 192: 92-116.
- [46] Pearce, D. 1984. "Rationalizable Strategic Behavior and the Problem of Perfection." *Econometrica* 52: 1029-1050.
- [47] Spence, M. 1977. "Consumer Misperceptions, Product Failure and Producer Liability." *The Review of Economic Studies* 44: 561-572.
- [48] Verrecchia, R. 1983. "Discretionary Disclosure." *Journal of Accounting and Economics* 5: 179-194.
- [49] Viscusi, W. K. 1978. "A Note on 'Lemons' Markets with Quality Certification." *The Bell Journal of Economics* 9: 277-279.